SKRIPSI

MODEL REGRESI SEMIPARAMETRIK SPLINE UNTUK DATA LONGITUDINAL PADA KASUS PENDERITA DEMAM BERDARAH DENGUE DI KOTA MAKASSAR

MUSTATI'ATUL WAIDAH MAKSUM

1511141001

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI MAKASSAR

2019

SKRIPSI

MODEL REGRESI SEMIPARAMETRIK SPLINE UNTUK DATA LONGITUDINAL PADA KASUS PENDERITA DEMAM BERDARAH DENGUE DI KOTA MAKASSAR

Diajukan kepada Program Studi Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam untuk memenuhi salah satu syarat memperoleh gelar Sarjana Sains.

MUSTATI'ATUL WAIDAH MAKSUM

1511141001

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI MAKASSAR

2019

PENGESAHAN SKRIPSI

Skripsi atas nama Mustati'atul Waidah Maksum, NIM : 1511141001 dengan judul Model Regresi Semiparametrik Spline untuk Data Longitudinal Pada Kasus Penderita Demam Berdarah Dengue Di Kota Makassar, diterima oleh Panitia Ujian Skripsi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Makassar, dengan SK. No. 1587/UN36.1/PP/2019, Tanggal 24 Mei 2019 untuk memenuhi sebagian persyaratan guna memperoleh gelar Sarjana Matematika pada Jurusan Matematika pada Hari Jumat, Tanggal 21 Juni 2019.

Disahkan Oleh:
Dekan FMIPA UNM Makassar

Drs. Suwardi Annas, M.Si., Ph.
NIP. 19691231 199403 1 110

Panitia Ujian:

1. Ketua Ujian : *Dr. Awi, M.Si.* (.................................)
2. Sekretaris : *Sutamrin, S.Si, M.Pd* (.................................)
3. Pembimbing I : *Prof. Syafruddin Side, M.Si., Ph.D.* (.................................)
4. Pembimbing II : *Wahidah Sanusi, S.Si., M.Si.,Ph.D.* (.................................)
5. Penguji I : *Muhammad Abdy, S.Si., M.Si, Ph.D.* (.................................)
6. Penguji II : *Ahmad Zaki, S.Si., M.Si.* (.................................)

PERNYATAAN KEASLIAN

Saya bertanda tangan di bawah ini menyatakan bahwa skripsi ini adalah hasil karya sendiri, dan semua sumber baik yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar. Bila dikemudian hari ternyata pernyataan saya terbukti tidak benar, maka saya bersedia menerima sanksi yang telah ditetapkan oleh FMIPA UNM Makassar.

Yang membuat pernyataan

Nama : Mustati'atul Waidah Maksum

NIM : 1511141001

Tanggal : Juni 2019

PERSETUJUAN PUBLIKASI UNTUK KEPENTINGAN AKADEMIK

Sebagai civitas akademi Universitas Negeri Makassar, saya bertanda tangan di bawah ini:

Nama	: Mustati'atul Waidah Maksum
Nim	: 1511141001
Program Studi	: Matematika
Jurusan	: Matematika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam

demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Negeri Makassar **Hak Bebas Royalti Non-eksklusif *(Non-exclusive Royalti-Free Right)*** atas skripsi saya yang berjudul: ***Model Semiparametrik Spline Untuk Data Longitudinal Pada Kasus Penderita Demam Berdarah Dengue di Kota Makassar*** beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-eksklusif ini Universitas Negeri Makassar berhak menyimpan, mengalih media/formatkan, mengelolah dalam bentuk pangkalan data *(database)*, merawat, dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis, pencipta dan pemilik hak cipta serta tidak dikomersilkan.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Makassar
Pada tanggal : Juni 2019

Menyetujui

Pembimbing I

Prof. Dr. Syarifuddin Side, M.Si
NIP. 19720202 199702 1 002

Yang Menyatakan

Mustati'atul Waidah Maksum
NIM. 1511141001

MUTIARA HIKMAH DAN PERSEMBAHAN

“Kamu tidak harus hebat untuk memulai, tapi kamu harus memulai untuk menjadi hebat”

(Zig Ziglar)

“Satu hal yang sangat buruk, jika seseorang berhenti di tempat dimana ia masih bisa berlanjut”

(Quraish Shihab)

Karya sederhana ini kupersembahkan kepada Ibu dan Ayah atas do’a, nasihat, pengorbanan, penantian serta sujud-sujud panjangnya yang penuh rasa cinta dan kasih sayang demi menyaksikan kesuksesan penulis dalam menggapai cita-cita dan kebahagiaan.

ABSTRAK

Mustati'atul Waidah Maksum, 2019. Model Regresi Semiparametrik *Spline* Untuk Data Longitudinal Pada Kasus Penderita Demam Berdarah Dengue di Kota Makassar. Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Makassar (dibimbing oleh Prof. Dr. Syafruddin Side, M.Si. dan Dr. Hj.Wahidah Sanusi,S.Si., M.Si.).

Regresi semiparametrik merupakan model regresi yang memuat komponen parametrik dan komponen nonparametrik dalam suatu model. Pada penelitian ini digunakan model regresi semiparametrik *spline* untuk data longitudinal dengan studi kasus penderita Demam Berdarah Dengue (DBD) di Rumah Sakit Unhas Makassar periode bula Januari sampai bulan Maret 2018. Estimasi model regresi terbaik didapat dari pemilihan titik knot optimal dengan melihat nilai *Generalized Cross Validation* (GCV) dan *Mean Square Error* (MSE) yang minimum. Komponen parametrik pada penelitian ini adalah hemoglobin (g/dL) dan umur (tahun), suhu tubuh (℃), trombosit ($\times 10^3 \mu L$) sebagai komponen nonparametrik dengan nilai GCV minimum sebesar 221,67745153 dicapai pada titik knot yaitu 14,552; 14,987; dan 15,096; nilai MSE sebesar 199,1032; dan nilai koefisien determinasi sebesar 75,3% yang diperoleh dari model regresi semiparametrik *spline* linear dengan tiga titik knot.

Kata kunci: regresi semiparametrik, *spline*, knot, GCV

ABSTRACT

Mustati'atul Waidah Maksum, 2019. Spline Semiparametric Regression Model for Longitudinal Data of Patient with Dengue Hemorrhagic Fever (DHF) of Makassar. Math Department, Faculty of Math and Natural Science, Makassar State University (supervised by Prof. Dr. Syafruddin Side, M.Si. dan Dr. Hj.Wahidah Sanusi,S.Si., M.Si).

Semiparametric regression is a regression model that includes parametric components and nonparametric components in a model. The regression model in this research is *spline* semiparametric regression with case studies of patients with Dengue Hemorrahagic Fever (DHF) at Unhas Makassar Hospital during the period of January to March 2018. The best regression model estimation is obtained from the selection of optimal knot which has minimum *Generalized Cross Validation* (GCV) and *Mean Square Error* (MSE). Parametric component in this research is hemoglobin (g/dL) and age (years), body temperature (℃), platelets $(\times 10^3 \mu L)$ as a nonparametric components. The minimum value of GCV is 221,67745153 achieved at the point 14,552; 14,987; dan 15,096 knot; MSE value of 199,1032; dan the value of coefficient determination is 75,3% obtained from semiparametric regression model linear spline with third point of knots.

Keywords: semiparametric regression, *spline*, knot, GCV

KATA PENGANTAR

Assalamu 'Alaikum Warahmatullahi Wabarakatuh.

Syukur Alhamdulillah penulis panjatkan kehadirat Allah SWT, atas berkat rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul "Model Regresi Semiparametrik Spline Untuk Data Longitudinal Pada Kasus Penderita Demam Berdarah Dengue di Kota Makassar", sebagai salah satu syarat menyelesaikan studi di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Makassar. Shalawat serta salam semoga senantiasa tercurahkan kepada nabi Muhammad SAW, para keluarga, sahabat, dan orang-orang yang senantiasa istiqamah di atas ajarannya hingga akhir zaman.

Sebagai seorang manusia biasa, penulis menyadari bahwa dalam penulisan skripsi ini tidak sedikit hambatan dan tantangan yang penulis hadapi. Akan tetapi dengan pertolongan Allah SWT., yang datang melalui dukungan dari berbagai pihak, baik secara langsung maupun tidak langsung sehingga semuanya dapat teratasi.

Penghargaan dan rasa terima kasih yang setinggi-tingginya penulis haturkan kepada:

1. Keluarga Tercinta, Ayahanda Drs. Maksum, Ibunda Cahaya Djunaid, SH., S.Pd, Kakanda Rifa'atul Mahmudah Maksum, Kakanda Miftahusshalihah Maksum dan Gunawan, Kakanda Nur Musyarafah Maksum dan Rusmin Aziz, Kakanda Khairun Nisa Maksum, Adinda Muhammad Ma'azim Maksum, dan

Keponakan tercinta Ghaisan Athaillah Gunawan dan Naufal Al Karni Rusmin atas motivasi, semangat, kasih sayang, serta fasilitas yang diberikan selama penyusunan skripsi ini.

2. Ayahanda Prof. Dr. H. Husain Syam, M.TP. selaku Rektor Universitas Negeri Makassar.
3. Ayahanda Drs. Suwardi Annas, M.Si., Ph.D selaku Dekan Fakultas Matematika dan llmu Pengetahuan Alam UNM.
4. Ayahanda Dr. Awi Dassa, M.Si. selaku Ketua Jurusan Matematika FMIPA UNM Makassar.
5. Ibu Dr. Hj. Wahidah Sanusi, S.Si., M.Si. selaku Ketua Program Studi Matematika, Jurusan Matematika FMIPA UNM Makassar dan sebagai Pembimbing II dalam penyusunan tugas akhir ini atas arahan dan kesediaannya dalam membimbing penulis.
6. Ayahanda Prof. Dr. Syafruddin Side, M.Si. selaku Pembimbing I dalam Penyusunan tugas akhir ini atas arahan dan kesediaannya dalam membimbing penulis.
7. Ayahanda Muhammad Abdy, S.Si., M.Si., Ph.D. dan Ayahanda Ahmad Zaki, S.Si., M.Si. selaku Penguji I dan Penguji II atas arahan dalam penyusunan tugas akhir ini.
8. Bapak/Ibu dosen Jurusan Matematika FMIPA UNM yang telah menyalurkan ilmunya secara ikhlas serta mendidik penulis.

9. Pihak Rumah Sakit Unhas Makassar yang telah memberikan kesempatan pada penulis untuk melakukan pengamatan dan pengambilan data untuk keperluan penelitian
10. Saudara sepupu Naiman Malik yang selalu memberikan bantuan selama pengambilan data.
11. Teman-teman yang telah membantu saya dalam mengolah data, Titi Kurnianti Hr dan Aditio Putra.
12. Teman yang menemani dari Maba hingga sekarang, Rasmini.
13. Teman-teman selalu menemani dan memberikan bantuan selama proses penyusunan skripsi ini, Usni Hamama, Citra Suci Said, dan Beby Fitriani.
14. Teman-teman seperjuangan “Titik Garis” (Matematika 2015), Musda, Nisa, Utami, Nurmah, Nensi, Arman, Hafila, Ade, Alam, Rifky, Adi, Amni, Rani, Risna, Nadia, Lana, Farid, Rahmat, Ray, Aswan, Fadlan, Syafri, Tyo, Ilo, dan Ika.
15. Teman-teman yang selalu memberikan dukungan, masukan, serta pengalaman mereka yang lebih dulu menyelesaikan studi nya, Miftahul Khaera, Sutarni, Ummu Shalwa, Rizkiya Aprianti, dan Miftakhaeria.
16. Teman-teman seperjuangan KKN-PPM Kab. Pinrang 2018, Aldi, Wana, Ana, Putri, Arik, Anggie, Ade Aviska, Fiska.

Serta orang-orang yang telah berjasa kepada penulis yang tidak dapat dituliskan oleh penulis. Penulis berharap semoga bantuan yang telah diberikan mendapatkan balasan dari Allah, sebagai amal jariyah dan pahala yang berlipat ganda di sisi-Nya.

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

Lampiran A

Lampiran B

BAB I

PENDAHULUAN

A. Latar Belakang

Data longitudinal sangat umum digunakan baik dalam studi observasional maupun studi eksperimental. Data longitudinal merupakan salah satu bentuk data berkorelasi. Dalam studi longitudinal dimungkinkan untuk mempelajari perubahan respons antar waktu beserta faktor yang mempengaruhi perubahan tersebut, baik pada level populasi maupun level individu. Penentuan pilihan dimensi waktu sangat tergantung pada pertanyaan penelitian yang ingin dijawab atau tujuan penelitian yang ingin dicapai (Poerwanto, 2014).

Inferensi pada model data longitudinal didasarkan pada data individu, dengan asumsi masing-masing individu independen, tetapi dengan memperhatikan bahwa observasi berulang untuk tiap-tiap individu tidak independen. Informasi yang diambil dari tiap individu atau variabel penelitian dalam penelitian longitudinal biasanya lebih dari satu variabel, yang dapat dikategorikan sebagai variabel respons (variabel dependen) dan variabel perjelas (variabel independen), (Danardono, 2018). Data longitudinal dicirikan oleh fakta bahwa pengamatan berulang dalam subyek yang sama cenderung, sehingga model-model untuk analisis data longitudinal harus mengenali hubungan antara pengamatan berkala dalam subyek yang sama (Notobroto, 2013).

Menurut Wu dan Zhang (2006) kelebihan menggunakan data longitudinal adalah dapat mengetahui perubahan yang terjadi pada individu, tidak membutuhkan subjek yang banyak karna pengamatannya berulang dan juga estimasinya lebih efisien karna dilakukan setiap pengamatan (Yunianti, 2015).

Salah satu tujuan menggunakan data longitudinal adalah untuk meneliti apakah ada pengaruh variabel perjelas terhadap variabel respons, termasuk meneliti pengaruh variabel perjelas terhadap besarnya perubahan variabel respons, sehingga pada dasarnya analisis data longitudinal adalah regresi pada data longitudinal (Danardono, 2018).

Analisis tentang pemodelan data longitudinal dilakukan dengan regresi semiparametrik spline. Regresi semiparametrik adalah gabungan antara regresi parametrik dan regresi nonparametrik. Pada pendekatan regresi parametrik diasumsikan bahwa bentuk kurva regresi diketahui berdasarkan informasi sebelumnya dan teori, ataupun pengalaman masa lalu. Akibatnya estimator kurva regresi diperoleh dengan mengestimasi parameternya (Budiantara, 2010). Pendekatan regresi nonparametrik tidak memberikan asumsi bentuk kurva tertentu ataupun tidak ada informasi mengenai bentuk kurva regresi. Kurva regresi dapat diasumsikan mulus atau *smooth*, sehingga regresi nonparametrik memiliki fleksibilitas yang tinggi karena data diharapkan mencari sendiri bentuk estimasi kurva regresinya tanpa dipengaruhi oleh faktor subyektifitas peneliti (Utami, 2014). Jika variabel respon diketahui hubungannya dengan salah satu variabel prediktor, tetapi dengan variabel prediktor yang lain tidak diketahui bentuk pola hubungannya. (Utami, 2014).

Pendekatan regresi nonparametrik telah banyak dikembangkan antara lain menggunakan spline, kernel, polinomial lokal, wavelet, dan fourier. Salah satu model regresi dengan pendekatan nonparametrik yang sangat sering digunakan untuk melakukan estimasi terhadap kurva regresi adalah regresi spline. (Adawiyah, 2018)

Spline adalah salah satu jenis *piecewise polynomial*. Maksud *piecewise polynomial* adalah polinomial yang memiliki sifat tersegmen atau sifat terpotong-potong. Model polinomial dengan sifat terpotong-potong menyebabkan *spline* memiliki fleksibilitas yang lebih tinggi dari model polinomial biasa, sehingga menyebabkan regresi *spline* dapat menyesuaikan diri secara lebih efektif terhadap karakteristik lokal suatu fungsi data atau dengan kata lain regresi *spline* dapat menghasilkan suatu fungsi regresi yang sesuai dengan data. (Sumarjaya, 2017)

Regresi *spline* adalah model regresi dengan kurva regresinya (fungsi regresinya) berupa fungsi *spline*. Pendekatan regresi *spline* ini tidak terikat akan asumsi bentuk kurva tertentu dan cenderung mencari sendiri estimasinya kemanapun pola data tersebut bergerak sehingga model yang diperoleh sesuai dengan bentuk data (Budiantara, 2011). Selain itu, metode *spline* ini sangat baik dalam memodelkan data yang polanya berubah-ubah pada sub interval tertentu. Pendekatan *spline* juga mempunyai keunggulan dalam mengatasi pola data yang menunjukkan naik atau turun yang tajam dengan bantuan titik-titik *knot,* serta kurva yang dihasilkan relatif mulus. Titik knot merupakan perpaduan bersama yang menunjukkan pola perilaku fungsi spline pada selang yang berbeda (Adawiyah,2018).

Regresi spline menjadi pendekatan popular untuk pemulusan data. Langkah awal yang dilakukan dalam regresi spline untuk mencari model terbaik adalah menentukan knot dengan nilai *Generalized Cross Validation* (GCV) yang minimum. Selain melihat GCV yang minimum, kriteria lain yang dapat digunakan adalah dengan melihat nilai *Mean Square Error* (MSE). Knot dapat diartikan sebagai suatu titik fokus dalam fungsi spline sehingga kurva yang dibentuk dapat terbagi pada titik tersebut. Dalam fungsi spline, memungkinkan digunakan berbagai macam orde, sehingga dapat dibentuk regresi spline linear, regresi spline kuadratik, dan seterusnya. Orde dalam fungsi spline diartikan sebagai pangkat terbesar dalam fungsi spline. Spline mempunyai kelemahan pada saat orde spline tinggi, maka dari itu penulis membatasi penelitiannya sampai orde 1. (Adawiyah, 2018).

Apabila pada sebuah model regresi terdapat komponen model yang diestimasi secara parametrik dan komponen lain menggunakan pendekatan nonparametrik maka terbentuklah model regresi semiparametrik. Keberadaan dua komponen yang berbeda dalam regresi semiparametrik ini menjadikan pemakaian model ini menjadi luas dan secara teori berkembang pesat. Sebelumnya telah dilakukan penelitian mengenai pemodelan regresi nonparametrik pada data longitudinal oleh Utami, 2014 dalam penelitiannya hanya menggunakan regresi nonparametrik dalam memodelkan data longitudinal. Tetapi dalam penelitian ini, akan digunakan regresi semiparametrik spline dalam memodelkan data longitudinal. Hasil pemodelan regresi semiparametrik pada data longitudinal diperoleh dari data kasus pasien demam berdarah dengue.

B. Rumusan Masalah

Berdasarkan latar belakang di atas, permasalahan yang dirumuskan adalah

1. Bagaimana menentukan model regresi parametrik pada pemodelan data longitudinal ?
2. Bagaimana menentukan model regresi nonparametrik pada pemodelan data longitudinal ?
3. Bagaimana model regresi semiparametrik Spline untuk data longitudinal pada kasus Demam Berdarah Dengue (DBD) ?

C. Tujuan Penelitian

Dari rumusan masalah di atas, tujuan dari penulisan ini adalah untuk:

1. Mengetahui model regresi parametrik pada pemodelan data longitudinal
2. Mengetahui model regresi nonparametrik pada pemodelan data longitudinal
3. Mengetahui model regresi semiparametrik spline untuk data longitudinal pada kasus Demam Berdarah Dengue (DBD)

D. Manfaat Penelitian

Manfaat yang diperoleh dari penulisan ini adalah:

1. Memberikan pengetahuan dasar tentang alternatif model regresi semiparametrik spline serta menambah wawasan tentang analisis data longitudinal.

2. Sebagai masukan bagi pemerintah pusat maupun pemerintah daerah, khususnya Dinas Kesehatan, dalam rangka pengambilan kebijakan program peningkatan derajat kesehatan masyarakat
3. Sebagai tambahan referensi bagi Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Makassar.

BAB II

KAJIAN PUSTAKA

A. Analisis Regresi

Analisis regresi merupakan suatu studi yang digunakan untuk melilhat ketergantungan atau hubungan antara suatu variabel respons (variabel terikat) pada satu atau lebih variabel prediktor (variabel bebas). Analisis regresi dapat dilakukan dengan tiga pendekatan yaitu parametrik, semiparametrik, dan nonparametrik (Sumarjaya, 2017). Model persamaan regresi adalah sebagaimana persamaan (2.1).

$$y_i = f(x_i) + \varepsilon_i, i = 1, 2, \ldots, n \tag{2.1}$$

Dimana

y_i : variabel respon atau dependen

x_i : variabel prediktor atau independen

$f(x_i)$: kurva regresinya

ε_i : error (kesalahan) acak yang diasumsikan identik, independen dan berdistribusi normal dengan mean nol dan variansi σ^2 atau $N(0, \sigma^2)$.

Analisis regresi merupakan metode statistik yang digunakan untuk menyelidiki hubungan atau pengaruh antara suatu peubah dengan peubah lainnya. Peubah-peubah regresi yang berhubungan secara linear disebut sebagai regresi linear. Regresi linear yang menghubungkan satu peubah terikat dengan satu peubah bebas disebut regresi linear sederhana, sedangkan regresi linear yang menghubungkan satu peubah terikat dengan dua atau lebih peubah bebas disebut regresi linear berganda.

Tujuan dari analisis regresi adalah mendapatkan estimasi parameter yang sesuai dengan bentuk kurva regresi. Jika bentuk kurva regresi diketahui, maka dapat digunakan pendekatan parametrik, sedangkan jika bentuk kurva regresi tidak diketahui dan tidak terdapat informasi yang lengkap sebelumnya maka dapat menggunakan pendeketan nonparametrik. Pendekatan semiparametrik dapat digunakan jika pola hubungan antara variabel prediktor dan respon merupakan kombinasi antara parametrik dan nonparametrik (Budiantara, 2014).

Kurva regresi hanya diasumsikan mulus (*smooth*) dalam arti termuat di dalam suatu ruang fungsi tertentu. Dengan demikian, pendekatan regresi nonparametrik memiliki fleksibelitas yang tinggi. Regresi semiparametrik merupakan gabungan antara regresi parametrik dan regresi nonparametrik. Model regresi semiparametrik mulai berkembang pesat sejak Wahba pada tahun 1990 mempublikasikan suatu model hubungan semiparametrik dari penggunaan listrik pada negara bagian di Amerika Serikat, yang berpola parametrik linear dengan pendapatan dan berpola nonparametrik dengan temperature (Adawiyah, 2018).

B. Regresi Parametrik

Regresi parametrik merupakan metode statistika yang digunakkan untuk mengetahui pola hubungan antara variabel prediktor dengan variabel respon dengan asumsi bahwa telah diketahui bentuk fungsi regresinya. Hubungan antara variabel respon dan variabel prediktor dalam model dapat terjadi dengan fungsi inier maupun nonlinier dalam parameter (Gusti, 2011).

Secara umum model regresi parametrik denggan variabel prediktor ke-k adalah

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i,\ i = 1,2,\dots,n \tag{2.2}$$

dimana

y_i : variabel respon

$x_{i1}, x_{i2}, \dots, x_{ik}$: variabel prediktor

$\beta_0, \beta_1, \dots, \beta_k$: parameter yang tidak diketahui

ε_i :error acak yang diasumsikan identik, independen dan berdistribusi normal dengan mean nol dan varians σ^2

Model regresi linier berganda pada persamaan (2.2) dapat ditulis dalam bentuk matriks sebagaimana persamaan (2.3).

$$\tilde{y} = \boldsymbol{X}\tilde{\beta} + \tilde{\varepsilon} \tag{2.3}$$

dengan:

$$\tilde{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{2.4}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \tag{2.5}$$

$$\tilde{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \tag{2.6}$$

dan $\tilde{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$ (2.7)

Untuk memperoleh estimator dari parameter biasanya digunakan metode *Ordinary Least Square* atau *Maximum Likelihood* (Adawiyah, 2018).

Dengan $\boldsymbol{\beta}$ vektor parameter berukuran $(p+1)\,x\,1$ yang akan diestimasi dari data, $\boldsymbol{Y}$ vektor observasi berukuran $n\,x\,1$, $\boldsymbol{X}$ matriks data berukuran $n\,x\,(p+1)$ yang diasumsikan mempunyai rank kolom penuh *(full rank)* dan $\boldsymbol{\varepsilon}$ vektor sesatan random dengan mean 0. Dengan metode kuadrat terkecil (*least square)* dapat diperoleh estimator untuk parameter β sebagaimana persamaan (2.9).

$$\tilde{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\tilde{y} \qquad (2.8)$$

Apabila diasumsikan $\varepsilon \sim IIDN\ (0, \sigma^2 I_n)$ maka

$$\boldsymbol{\hat{\beta}} \sim \boldsymbol{N}(\boldsymbol{\beta}, (\boldsymbol{X^T X})^{-1}\boldsymbol{\sigma^2}) \qquad (2.9)$$

Pendekatan pendugaan yang paling sering digunakan adalah pendekatan parametrik. Asumsi yang mendasari pendekatan ini adalah kurva regresi dapat diwakili oleh suatu model parametrik. Dalam regresi parametrik, diasumsikan bahwa bentuk kurva regresi diketahui berdasarkan teori, informasi sebelumnya, atau sumber-sumber lain yang dapat memberi pengetahuan secara terperinci (Adawiyah, 2018).

Pendekatan parametrik mengasumsikan bentuk fungsi regresi tertentu dan distribusi galatnya harus memenuhi asumsi tertentu seperti normalitas, homoskedastisitas, tidak terjadi autokorelasi dan multikoliniearitas. Asumsi-asumsi tersebut sangat berpengaruh terhadap model regresi. Dalam model regresi parametrik, terdapat dua model yaitu model linear dan non linear (Adawiyah, 2018).

C. Regresi Nonparametrik

Regresi nonparametrik digunakan apabila bentuk pola hubungan antara variabel respon dengan variabel prediktor tidak diketahui bentuk kurva regresinya. Dalam regresi nonparametrik kurva regresi hanya diasumsikan mulus (*smooth*) dalam arti termuat dalam suatu ruang fungsi tertentu sehingga mempunyai sifat fleksibilitas yang tinggi (Gusti, 2011).

Secara umum model regresi nonparametrik dapat dituliskan sebagai berikut:

$$y_i = f(x_i) + \varepsilon_i, i = 1,2,3, \dots, n \tag{2.10}$$

dimana

y_i : variabel respon

$f(x_i)$: fungsi *smooth* yang tidak diketahui ke-i

ε_i :error acak yang diasumsikan identik, independen, dan berdistribusi normal dengan mean nol dan varians σ^2

Jika diberikan matriks pada persamaan (2.11) – (2.13), maka model regresi pada persamaan (2.11) dapat dituliskan sebagaimana bentuk matriks pada persamaan (2.14)

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{2.11}$$

$$f(X) = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} \tag{2.12}$$

$$\text{dan } \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{2.13}$$

$$y = f(X) + \varepsilon \tag{2.14}$$

Pendekatan nonparametrik digunakan untuk mengestimasi kurva regresi karena model tidak ditentukan terlebih dahulu seperti pada regresi parametrik. Salah satu pendekatan nonparametrik yang bisa dilakukan adalah dengan fungsi *spline*. (Laome, 2009)

D. Regresi Semiparametrik

Model regresi semiparametrik merupakan gabungan dari model regresi parametrik dan regresi nonparametrik (Gusti, 2011). Adapun model regresi semiparametrik adalah

$$Y_i = X_i\beta + f(t_i) + \varepsilon_i, i = 1,2,3, \dots, n \tag{2.15}$$

Dimana

Y_i : nilai variabel respon dalam amatan ke-i

X_i : peubah bebas atau variabel prediktor yang berhubungan secara parametrik dengan variabel respon Y_i

β : $(\beta_0, \beta_1, \dots, \beta_h)$ merupakan parameter koefisien regresi

$f(t_i)$: fungsi *smooth* yang tidak diketahui bentuk polanya terhadap variabel respon (berhubungan secara nonparametrik dengan variabel respon)

ε_i :error acak yang diasumsikan identik, independen, dan berdistribusi normal dengan mean nol dan varians σ^2

E. Regresi Spline

Spline merupakan potongan polinomial (piecewise polynomial) tersegmen yang memiliki sifat fleksibilitas. Titik perpaduan bersama dari potongan-potongan tersebut atau titik yang menunjukkan terjadinya perubahan-perubahan perilaku kurva pada interval-interval yang berbeda disebut knot (Adawiyah, 2018).

Adapun model dari regresi spline adalah sebagaimana persamaan (2.16) (Adawiyah, 2018).

$$f(t_i) = \sum_{j=0}^{m} \beta_j t_i^j + \sum_{l=1}^{p} \beta_{(m+l)} (t_i - k_l)^m \tag{2.16}$$

dengan $(t_i - k_l)^m = \begin{cases} (t_i - k_l)^m; & t_i \geq k_l \\ 0 & ; \ t_i < k_l \end{cases}$

dimana

$f(t_i)$	: fungsi regresi spline
$k_1, k_2, \dots, k_k$	: titik knot
t	: variabel prediktor
β	: konstanta

Taksiran kurva $f(t)$ adalah $\hat{f}_\lambda(t)$ yakni penaksir kurva yang mulus, diperoleh melalui model regresi polinomial. Dengan mempertimbangkan sifat-sifat fungsi spline, yang merupakan modifikasi dari regresi polynomial, maka untuk mendapatkan model taksiran dari kurva digunakan regresi spline.

Dengan demikian bentuk umum regresi nonparametrik dengan pendekatan spline orde ke-m adalah sebagaimana persamaan (2.17).

$$y_i = \beta_0 + \sum_{j=1}^{m} \beta_j t_i^j + \sum_{l=1}^{p} \beta_{(m+l)} (t_i - k_l)^m + \varepsilon_i \tag{2.17}$$

$y_i = \hat{y} + \varepsilon_i$

Selanjutnya model regresi spline dapat ditulis sebagaimana persamaan (2.18).

$$y_i = \beta_1 t^1 + \cdots + \beta_m t^m + \beta_1 (t_i - k_1)^m + \cdots + \beta_l (t_i - k_l)^m + \varepsilon_i \tag{2.18}$$

Dengan menggunakan data sebanyak n, maka bentuk matriks dari persamaan (2.18) ditulis sebagai berikut:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} t_1 & \dots & t_1^m & (t_1 - k_1)^m & \cdots & (t_1 - k_l)^m \\ t_2 & \dots & t_2^m & (t_2 - k_1)^m & \dots & (t_2 - k_l)^m \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ t_n & \dots & t_n^m & (t_n - k_1)^m & \dots & (t_n - k_l)^m \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \\ \beta_1 \\ \vdots \\ \beta_l \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{2.19}$$

F. Pendugaan Parameter Generalized Estimating Equation (GEE)

Generalized Estimating Equation (GEE) merupakan perkembangan dari *Generalized Linear Model* (GLM) yang diperkenalkan oleh Liang dan Zeger (1986) yang digunakan untuk menduga parameter model berdasarkan data yang mengandung autokorelsi dan data yang tidak menyebar normal (Handayanti, 2015). Fungsi penghubung GEE adalah sebagai berikut :

$$g(\mu_{ij}) = X_i \beta \tag{2.20}$$

Penduga parameter β dapat diperoleh dengan menyelesaikan persamaan berikut (Danardono, 2015) :

$$\sum_{i=1}^{n} D_i^T V_i^{-1} (Y_i - \mu_i) = 0 \tag{2.21}$$

Dengan $D_i = \frac{\partial \mu_i}{\partial \beta}$. Matriks D_i disebut sebagai matriks derivatif, yaitu matriks yang berisi turunan μ_i terhadap komponen β. Matriks ini mentransformasikan unit asal berupa Y_{ij} (dan μ_i) menjadi unit pada skala $g(\mu_{ij})$. Skala pada unit fungsi penghubung $g(\mu_{ij})$ ini dapat digunakan untuk memberi interpretasi pada nilai β. Sedangkan V_j merupakan matriks ragam-peragam Y_{ij} yang berukuran $n_j \times n_j$ pada subyek ke-i, yakni :

$$V_j = \left[\emptyset A_j^{1/2} R(\alpha) A_j^{1/2}\right] \qquad (2.22)$$

Dimana φ adalah parameter dispersi yang diduga dengan :

$$\hat{\emptyset} = \frac{1}{n-p} \sum_{i=1}^{n} \sum_{j=1}^{p} e_{ij}^2 \qquad (2.23)$$

Dimana n merupakan ukuran contoh, p adalah banyaknya parameter dan e_{ij} adalah galat *Pearson*

$$e_{ij} = \frac{Y_{ij} - \mu_{ij}}{\sqrt{var(\mu_{ij})}} \qquad (2.24)$$

Dan $R(\alpha)$ adalah matriks korelasi berukuran $n_i \times n_i$ yang berisikan korelasi antar respon pada i subyek, untuk n data berulang

Pemilihan struktur korelasi bermanfaat untuk mendapatkan struktur korelasi yang sesuai dengan data. Korelasi dibentuk dalam sebuah matriks korelasi $R(\alpha)$ berukuran $n_j \times n_j$, dimana struktur korelasi yang tidak diketahui dan harus diduga (Handayanti, 2015).

Sedangkan pada fungsi mulus, Green dan Silverman memperkenalkan fungsi mulus untuk beberapa kondisi, misalnya pada regresi nonparametrik dan regresi semiparametrik untuk data independen kontinu, nonparametrik dan semiparametrik dengan pendekatan generalized linear models untuk data independent, dan quasi-likelihood untuk data independen. Ketiganya juga dapat digunakan pada data kontinu yang berkorelasi. Untuk pendekatan quasi-likelihood, hasil yang penting adalah solusi dari fungsi f untuk regresi nonparametrik dan parameter β dalam regresi semiparametrik dengan memaksimumkan "penalized quasi-likelihood" :

$$\Pi = \Phi - \frac{1}{2}\lambda \int_a^b [f''(t)]^2 dt \qquad (2.25)$$

Dimana $\Phi = \sum_{i=1}^{n} D_i^T V_i^{-1}(Y_i - \mu_i)$, $\lambda > 0$ adalah faktor penalized dan $\int_a^b [f''(t)]^2 dt = f^T G f$ dimana G adalah matriks simetris (Ibrahim, N.A., dan Suliadi, 2009).

G. Pemilihan Titik Knot Optimal

Titik knot merupakan titik perpaduan bersama yang memperlihatkan terjadinya perubahan perilaku dari fungsi *spline* pada interval-interval yang berbeda sehingga kurva yang terbentuk tersegmen pada titik tersebut. pada penetuan model regresi *spline* dapat dilakukan dengan melihat nilai *Generalized Cross Validation* (GCV) yang minimum.. Adapun rumus untuk menghitung GCV adalah sebagai berikut (Yani, 2017) :

$$GCV(K) = \frac{MSE(K)}{(n^{-1} tr[I - A(K)])^2} \qquad (2.26)$$

Dengan MSE(K) = $n^{-1}\sum_{i=1}^{n}[Y_i - \hat{f}(t_i)]^2$, K adalah titik knot $(K_1, K_2, K_3, \dots, K_n), \hat{f}(t_i) = t(t^1t)^{-1}t'Y$, n adalah jumlah data, I adalah matriks identitas, $A(K) = t(t^1t)^{-1}t'$ dan $\hat{Y}_i = A(K)Y$.

H. Pengujian Parameter Model

Pengujian parameter dalam model regresi bertujuan untuk mengetahui apakah parameter tersebut telah menunjukkan hubungan yang nyata antara variabel prediktor dan variabel respon. Selain itu juga untuk mengetahui kelayakan parameter dalam menerangkan model (Gusti, 2011).

1) Uji Simultan (Uji F)

   Uji simultan digunakan untuk memeriksa signifikansi koefisien regresi secara bersama-sama

   a. Hipotesis Pengujian

      $H_0: \beta_j = 0, \ \ \forall j = 0,1,2, \dots, n$

      H_1: minimal terdapat satu $\beta_j \neq 0$, untuk suatu $j = 0,1,2, \dots, n$

   b. Statistik Uji

$$F = \frac{KT_{Reg}}{S^2} \tag{2.27}$$

   c. Keputusan

      Jika $F_{hitung} > F_{tabel}$ maka tolak H_0dan terima H_1

2) Uji Individu atau Uji Parsial

   Uji parameter secara parsial (secara individu) menggunakan pendekatan GEE dapat dilihat berdasarkan nilai signifikansi atau *p-value* dari uji Wald.

Apabila $p - value$ kurang dari taraf nyata (α) maka parameter tersebut signifikan terhadap variabel respon.

I. Pengujian Asumsi Residual

Residual (*goodness of fit)* dari suatu model regresi harus memenuhi asumsi *IIDN* (0,σ^2) yaitu identik, independen, dan berdistribusi normal dengan mean nol dan variansi σ^2. Sehingga adapun tujuan dari melakukan pengujian asumsi residual ini yaitu untuk mengetahui apakah residual yang dihasilkan sudah memenuhi asumsi *IIDN* (0,σ^2) .

1) Uji Asumsi Homogenitas

Uji asumsi homogenitas yaitu uji asumsi identik terpenuhi ketika sebaran plot tidak membentuk suatu pola tertentu atau tersebar secara acak dan variansi residual bersifat homokedastisitas yang artinya semua memiliki nilai yang sama σ^2. Apabila sebaran plot tidak tersebar secara acak atau membentuk suatu pola tertentu maka mengindikasikan terjadi heteroskedastisitas. Selain dengan menggunakan metode grafis, identifikasi heteroskedastisitas juga dapat dilakukan dengan menggunakan uji Glejser. Uji Glejser mempertimbangkan regresi nilai $|e_i|$ terhadap variabel X yang dianggap berhubungan dekat dengan varians heteroskedastisitas α_i^2. Beberapa bentuk fungsional yang dianjurkan dalam regresi ini adalah (Yani, N.W.M.N, 2017) :

$$|e_i| = \gamma_0 + \gamma_1 X_i + \varepsilon_i, \qquad (2.28)$$

$$|e_i| = \gamma_0 + \gamma_1 \sqrt{X_i} + \varepsilon_i, \qquad (2.29)$$

$$|e_i| = \gamma_0 + \gamma_1 \left(\frac{1}{X_i}\right) + \varepsilon_i, \tag{2.30}$$

Adapun hipotesis dari uji Glejser adalah sebagai berikut :

$$H_0 : \gamma_1 = 0,$$

$$H_1 : \gamma_1 \neq 0$$

Dengan daerah penolakan yakni tolak H_0 apabila $p - value < \alpha$.

2) Uji Asumsi Independen

Uji asumsi independen dilakukan untuk mengetahui apakah terdapat korelasi pada residual. Asumsi residual independen dapat dilakukan dengan menggunakan uji d Durbin-Watson, adapun uji d Durbin-Watson dapat dirumuskan sebagai berikut (Yani, N.W.M.N, 2017) :

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} \tag{2.31}$$

Adapun hipotesis dari uji d Durbin-Watson adalah sebagai berikut :

$$H_0 : \rho = 0,$$

$$H_1 : \rho \neq 0$$

Dengan aturan keputusan sebagai berikut :

Tabel 2.1 Tabel aturan keputusan uji d Durbin-Watson

Hipotesis nol	Keputusan	Jika
Tidak ada autokorelasi positif	Tolak	$0 < d < d_L$
Tidak ada autokorelasi positif	Tidak ada keputusan	$d_L < d < d_u$
Tidak ada autokorelasi negatif	Tolak	$4 - d_L < d < 4$
Tidak ada autokorelasi negatif	Tidak ada keputusan	$4 - d_U \leq d < 4 - d_L$
Tidak ada autokorelasi positif atau negatif	Terima	$d_U \leq d < 4 - d_U$

3) Uji Asumsi Normalitas

Uji asumsi normalitas dilakukan untuk melihat apakah residual mengikuti distribusi normal atau tidak. Pengujian asumsi normalitas dapat dilakukan dengan menggunakan uji Anderson-Darling. Adapun uji Anderson-Darling dapat dirumuskan sebagai berikut (Yani, N.M.W.M, 2017) :

$$A^2 = -n - S \tag{2.32}$$

Dengan $S = \frac{1}{n}\sum_{i=1}^{n}[2i - 1][\log\left(F(Z_i)\right) + \log(1 - F(Z_{n+1-i}))]$ adalah simpangan baku data, $Z_i = \frac{X_i - \bar{X}}{\sigma}$ adalah data X_i yang distandarisasi, X_i adalah data ke-i yang telah diurutkan, $\bar{X}$ adalah rata-rata data, F(Z_i) adalah nilai fungsi distribusi kumulatif normal baku di Z_i, A^2 adalah statistik uji untuk metode Anderson-Darling, n adalah ukuran sampel, dan $F_0(X)$ adalah fungsi distribusi kumulatif teoritis. Dengan hipotesis yang digunakan untuk menguji normalitas residual adalah sebagai berikut :

$H_0: F_0(X) = F(X)$ (Residual berdistribusi $N(0, \sigma^2)$)

$H_1: F_0(X) \neq F(X)$ Residual tidak berdistribusi $N(0, \sigma^2)$)

Keputusan tolak H_0 apabila nilai A^2 lebih besar dari nilai kritisnya atau $p - value < \alpha$.

J. Data Longitudinal

Data longitudinal adalah data pengamatan berulang pada unit eksperimen, berbeda dengan data *cross section* yaitu data dari masing-masing individu diamati dalam sekali waktu. Ada beberapa keuntungan dari studi mengenai data longitudinal dibandingkan dengan data *cross section*. Pertama, studi longitudinal lebih *powerful* dari studi *cross section* untuk sejumlah subjek yang tetap. Dengan kata lain, untuk memperoleh kekuatan uji statistik yang sama, studi longitudinal membutuhkan subjek yang lebih sedikit. Kedua, dengan jumlah subjek yang sama, hasil pengukuran *error* menghasilkan penaksir efek perlakuan yang lebih efisien dari data *cross section*. Ketiga, data longitudinal mampu menyediakan informasi tentang perubahan individu, sedangkan data *cross section* tidak. (Laome, 2009).

Jika y_{ij}menyatakan pengamatan untuk subjek ke- i pada waktu ke- j , t_{ij} menyatakan variabel prediktor dan n adalah banyaknya subjek dan n_i menyatakan banyaknya ulangan pada subjek ke-i dalam kurun waktu berbeda maka diberikan data longitudinal $\{(t_{ij}, y_{ij}), j = 1,2, \dots, n_i; i = 1,2, \dots, n\}$ (Utami, 2014).

K. Demam Berdarah Dengue

Demam berdarah (DB) atau demam berdarah dengue (DBD) adalah penyakit febril akut yang ditemukan di daerah tropis, dengan penyebaran geografis yang mirip dengan malaria. Penyakit ini disebabkan oleh salah satu dari empat serotipe virus dari genus *Flavivirus*, famili *Flaviviridae*. Terdapat tiga faktor pemegang peran dalam penularan infeksi virus dengue yaitu manusia, virus, dan vector perantara (Sumarjaya, 2017). Demam berdarah disebarkan kepada manusia oleh nyamuk *Aedes aegypti*.

Penyakit ini dapat didiagnosis dengan melihat gejala awal yang muncul, seperti demam tinggi dan munculnya ruam. Gejala tersebut ada kesamaan dengan gejala dari penyakit malaria, leptospirosis, maupun demam tifoid, maka untuk mendapatkan ketepatan diagnosis yang lebih tinggi umumnya dilakukan berbagai uji laboraturium, seperti menghitung humlah antibodi terhadap virus dengue, dan perhitungan darah lengkap (hemoglobin, leukosit, hematokrit, dan trombosit). (Purhadi, 2012).

Demam berdarah umumnya lamanya sekitar enam atau tujuh hari dengan puncak demam yang lebih kecil terjadi pada akhir masa demam. Kadar trombosit yang normal berkisar antara 150-440/ribu/ml sedangkan ada penderita demam berdarah, trombosit dapat turun hingga dibawah 100 ribu/ml (Utami, 2014).

L. Penelitian Relevan

Peneliti (tahun)	Judul Penilitian	Kesimpulan	Keterkaitan
Tiani Wahyu Utami (2014)	Pemodelan Regresi Nonparametrik Pada Data Longitudinal Berdasarkan Estimator Polinomial Lokal Kernel	Bentuk estimasi model regresi nonparametrik pada data longitudinal berdasarkan estimator polinomial lokal kernel adalah $\hat{\eta}(t_{ij}) = x_{ij}^T(X^T K_h X)^{-1} X^T K_h y$	Pengembangan metode regresi nonparametrik menjadi regresi semiparametrik pada data longitudinal
Purhadi (2012)	Analisis *Survival* Faktor-Faktor yang Mempengaruhi Laju Kesembuhan Pasien Penderita Demam Berdarah Dengue (DBD) di RSU Haji Surabaya dengan Regresi *Cox*	Faktor-faktor yang mempengaruhi laju kesembuhan pasien adalah usia dan trombosit. Resiko sembuh pasien dengan usia satu tahun lebih tua akan lebih lama sembuh daripada usia yang lebih muda dan resiko untuk mencapai sembuh pasien dengan trombosit di bawah normal juga akan lebih lama sembuh daripada yang normal.	Faktor-faktor yang paling signifikan terhadap laju kesembuhan pasien penderita DBD
Ni Wayan Merry Nirmala Yani (2017)	Aplikasi Model Regresi Semiparametrik *Spline Truncated*	Estimasi model regresi semiparametrik *spline truncated* diperoleh $\hat{Y}_i = -0{,}55021 - 0{,}01136HCT_i + 0{,}01648(HCT_i - 39{,}6) + 0{,}28815S_i - 0{,}0475U_i - 0{,}01104PLT_i$ dengan nilai GCV minimum sebesar 0,03553, nilai MSE sebesar 0,02969, nilai koefisien determinasi sebesar 98,91% .	Penggunaan regresi semiparametrik *spline*

BAB III

METODE PENELITIAN

A. Jenis Penilitian

Penelitian yang dilakukan merupakan penelitian terapan dan tentang regresi semiparametrik menggunakan pendekatan spline, selain itu juga digunakan data penderita DBD dalam penerapan regresi semiparametrik dengan pendekatan spline tersebut.

B. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yaitu data penderita Demam Berdarah Dengue yang diperoleh dari Rumah Sakit UNHAS Makassar.

C. Lokasi dan Waktu Penelitian

Penelitian ini dilakukan di Perpustakaan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Makassar sebagai lokasi utama dalam pengumpulan literatur, serta Rumah Sakit UNHAS Makassar sebagai lokasi pengambilan data. Penelitian ini akan dilaksanakan pada Januari 2019.

D. Variabel Penelitian

Data yang digunakan adalah data penderita demam berdarah dengan variabel respon yaitu laju kesembuhan dan variabel prediktornya yaitu usia, kadar trombosit, hemoglobin, suhu tubuh.

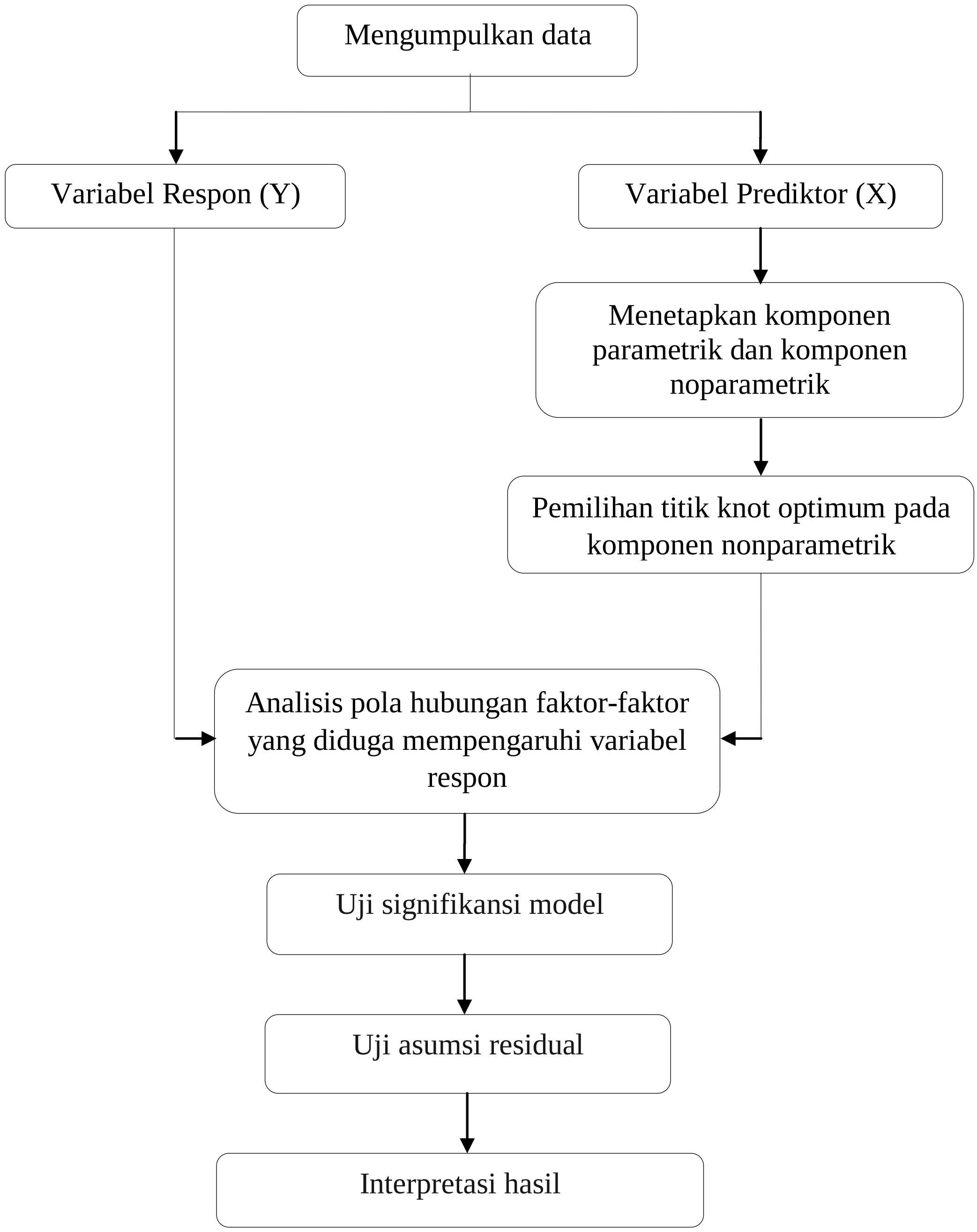
E. Prosedur Pelaksanaan Penelitian

Prosedur penelitian yang dilakukan untuk mencapai tujuan penelitian sebagaimana dijelaskan sebagai berikut:

1. Menetapkan komponen parametrik dan komponen nonparametrik berdasarkan data.
2. Untuk komponen Nonparametrik, dipilih titik *knot* optimal dengan menggunakan *Generalized Cross Validation (GCV)* yang paling minimum.
3. Memodelkan data kadar trombosit pasien demam berdarah dengan variabel prediktor yang telah ditetapkan dengan regresi Semiparametrik Spline dengan titik *knot* optimal
4. Menhitung nilai koefisien determinasi (R^2).
5. Menguji signifikansi parameter regresi semiparametrik spline secara serentak dengan uji F.
6. Melakukan uji parameter regresi semiparametrik spline secara parsial dengan uji t
7. Menguji asumsi residual IIDN dari model semiparametrik spline
8. Menginterpretasikan model regresi semiparametrik spline

F. Skema Penelitian

Adapun skema penelitian yang akan dilakukan dapat dilihat pada Gambar 3.1



Gambar 3.1 Skema Penelitian

BAB IV

HASIL DAN PEMBAHASAN

A. Estimasi Model Regresi Semiparametrik Spline dengan Pendekatan Generalized Estimating Equation (GEE)

Diberikan model regresi semiparametrik :

$$Y_{ij} = X_{ij}\beta + f(t_{ij}) + \varepsilon_{ij}, i = 1, \dots, n; j = 1, \dots, m \quad (1)$$

Dimana Y_{ij} adalah variabel respon kelompok ke-i untuk pengamatan ke-j, X_{ij} merupakan komponen parametrik dan $f(t_{ij})$ adalah komponen nonparametrik yang merupakan fungsi mulus yang tidak diketahui dan $\varepsilon_{ij} \sim NII(0, \sigma^2)$. Berdasarkan fungsi penghubung GEE, bentuk (1) diatas dapat dinyatakan dalam bentuk :

$$l(\mu_{ij}) = X_{ij}\beta + f(t_{ij}) \quad (2)$$

Dimana $l(.)$ adalah suatu fungsi penghubung dan $\mu_{ij} = E[Y_{ij}|X_{ij}]$. Parameter β akan diestimasi dengan menggunakan GEE dan fungsi mulus $f(t)$ akan diestimasi dengan memaksimumkan penalized quasi-likelihood.

Pendekatan GEE untuk mengestimasi β pada model regresi parametrik akan diuraikan sebagai berikut :

Misalkan $X_i = (X_{i1}, \dots, X_{in})^T$ adalah matriks $n \times p$ yang merupakan nilai dari variabel prediktor untuk subjek ke-i $(i = 1, \dots, n)$. Diberikan persamaan estimasi GEE sebagai berikut :

$$\sum_{i=1}^{n} D_i^T V_i^{-1}(Y_i - \mu_i) = 0 \qquad (3)$$

Dimana $\mu_i = E(Y_i)$ yang mempunyai komponen ke-j $\mu_{ij} = X_{ij}\hat{\beta}$, $D_i = \frac{\partial \mu_i}{\partial \beta}$, dan $V_i = \emptyset A_i^{\frac{1}{2}} R_i(\alpha) A_i^{1/2}$, dimana $R_i(\alpha)$ adalah matriks korelasi berukuran $n \times n$, untuk n data berulang untuk satu individu i, dan A_i adalah matriks diagonal berukuran $n \times n$ dengan $V(\mu_{ij})$ adalah elemen diagonalnya, $\emptyset$ adalah suatu parameter dispersi (penyebaran). Berdasarkan bentuk (3) diperoleh parameter $\hat{\beta}$:

$$\sum_{i=1}^{n} D_i^T V_i^{-1}(Y_i - \mu_i) = 0$$

$$\Leftrightarrow \sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \beta}\right)^T V_i^{-1}(Y_i - X_i\hat{\beta}) = 0$$

$$\Leftrightarrow \sum_{i=1}^{n} X_i^T V_i^{-1}(Y_i - X_i\hat{\beta}) = 0$$

$$\Leftrightarrow \sum_{i=1}^{n} X_i^T V_i^{-1} Y_i - \sum_{i=1}^{n} X_i^T V_i^{-1} X_i \hat{\beta} = 0$$

$$\Leftrightarrow \hat{\beta} = \left[\sum_{i=1}^{n} X_i^T V_i^{-1} X_i\right]^{-1} \left[\sum_{i=1}^{n} X_i^T V_i^{-1} Y_i\right]$$

Kemudian menggunakan estimasi parameter ke t untuk memperbaharui $\hat{\beta}$ dalam persamaan sebagai berikut :

$$\hat{\beta}^{t+1} = \hat{\beta}^{t} - \left[\sum_{i=1}^{n} X_i^T V_i^{-1} X_i\right]^{-1} \left[\sum_{i=1}^{n} X_i^T V_i^{-1} Y_i\right]$$

Selanjutnya, fungsi mulus $f(t)$ akan diestimasi dengan menggunakan GEE dengan memaksimumkan penalized quasi-likelihood. Definisi fungsi penalized quasi-likelihood adalah (Ibrahim, N.A., dan Suliadi, 2009) :

$$\Pi = \Phi - \frac{1}{2}\lambda \int_a^b [f''(t)]^2 dt \tag{4}$$

Dimana $\Phi = \sum_{i=1}^{n} D_i^T V_i^{-1}(Y_i - \mu_i)$, $\lambda > 0$ adalah faktor penalized dan $\int_a^b [f''(t)]^2 dt = f^T G f$ dimana G adalah matriks simetris (Ibrahim, N.A., dan Suliadi, 2009).

Misalkan $Y_i = (Y_{i1}, \dots, Y_{in})^T$ adalah vektor $n \times 1$ yang menyatakan variabel respons, $f = \left[f\left(t_{(1)}\right), f(t_{(2)}), \dots f(t_{(q)})\right]^T$. Selanjutnya, $f(.)$ akan diestimasi dengan memaksimumkan bentuk (4) sebagai berikut :

$$\frac{\partial \Pi}{\partial f} = \sum_{i=1}^{n} D_i^T V_i^{-1}(Y_i - \mu_i) - \frac{\partial}{\partial f}\left[\frac{1}{2}\lambda \int_a^b [f''(t)]^2 dt\right] \tag{5}$$

Perhatikan bahwa

$$\frac{\partial}{\partial f}\left[\frac{1}{2}\lambda \int_a^b [f''(t)]^2 dt\right] = \frac{1}{2}\lambda \frac{\partial}{\partial f} f^T G f$$

Misal

$$f = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_q \end{pmatrix}, \qquad f^T = \begin{pmatrix} f_1 & f_2 & \cdots & f_q \end{pmatrix}$$

$$G = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1m} \\ g_{21} & g_{22} & \cdots & g_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{nm} \end{pmatrix}$$

$$Z = f^T G f = (f_1 \quad f_2 \quad \cdots \quad f_q)\begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1m} \\ g_{21} & g_{22} & \cdots & g_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{nm} \end{pmatrix}\begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_q \end{pmatrix}$$

$$= \left(f_1 g_{11} + f_2 g_{21} + \cdots + f_q g_{n1} \quad f_1 g_{12} + f_2 g_{22} + \cdots + f_q g_{n2} \quad \cdots \quad f_1 g_{12} + f_2 g_{22} + \cdots + f_q g_{nm}\right)\begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_q \end{pmatrix}$$

$$= {f_1}^2 g_{11} + f_1 f_2 g_{21} + \cdots + f_1 f_q g_{n1} + f_2 f_1 g_{12} + {f_2}^2 g_{22} + \cdots + f_2 f_q g_{n2} + \cdots +$$

$$f_q f_1 g_{1m} + f_q f_2 g_{2m} + \cdots + {f_q}^2 g_{nm} \qquad (6)$$

Selanjutnya, persamaan (6) diatas diturunkan terhadap f_1, f_2 diperoleh

$$\frac{\partial Z}{\partial f} = \begin{pmatrix} \frac{\partial f}{\partial f_1} \\ \frac{\partial f}{\partial f_2} \\ \vdots \\ \frac{\partial f}{\partial f_q} \end{pmatrix} = \begin{pmatrix} 2f_1 g_{11} + f_2 g_{21} + \cdots + f_q g_{n1} + f_2 g_{12} + \cdots + f_q g_{1m} \\ f_1 g_{21} + f_1 g_{12} + 2f_2 g_{22} + \cdots + f_q g_{n2} + \cdots + f_q g_{2m} \\ \vdots \\ f_1 g_{n1} + f_2 g_{2n} + \cdots + f_1 g_{1m} + f_2 g_{2m} + \cdots + 2f_q g_{nm} \end{pmatrix}$$

Karena G matriks simetris, diperoleh

$$\frac{\partial Z}{\partial f} = \begin{pmatrix} 2f_1 g_{11} + 2f_2 g_{12} + \cdots + 2f_q g_{1m} \\ 2f_1 g_{21} + 2f_2 g_{22} + \cdots + 2f_q g_{2m} \\ \vdots \\ 2f_1 g_{n1} + 2f_2 g_{2n} + \cdots + 2f_q g_{nm} \end{pmatrix}$$

$$= 2\begin{pmatrix} f_1 g_{11} + f_2 g_{12} + \cdots + f_q g_{1m} \\ f_1 g_{21} + f_2 g_{22} + \cdots + f_q g_{2m} \\ \vdots \\ f_1 g_{n1} + f_2 g_{2n} + \cdots + f_q g_{nm} \end{pmatrix}$$

$$= 2\begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1m} \\ g_{21} & g_{22} & \cdots & g_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{nm} \end{pmatrix}\begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_q \end{pmatrix}$$

$$= 2Gf$$

Sehingga diperoleh

$$\frac{\partial}{\partial f}\left[\frac{1}{2}\lambda\int_a^b [f''(t)]^2 dt\right] = \frac{1}{2}\lambda\ \frac{\partial}{\partial f} f^T G f$$

$$= \frac{1}{2}\lambda\ 2Gf$$

$$= \lambda G f$$

Akibatnya, bentuk (5) dapat diselesaikan sebagai berikut :

$$\frac{\partial \Pi}{\partial f} = \sum_{i=1}^{n} D_i^T V_i^{-1}(Y_i - \mu_i) - \frac{\partial}{\partial f}\left[\frac{1}{2}\lambda\int_a^b [f''(t)]^2 dt\right]$$

$$= \sum_{i=1}^{n} D_i^T V_i^{-1}(Y_i - \mu_i) - \lambda G \hat{f}$$

Selanjutnya, dari persamaan diatas diperoleh fungsi mulus $f(.)$

$$\sum_{i=1}^{n} D_i^T V_i^{-1}(Y_i - \mu_i) - \lambda G \hat{f} = 0$$

$$\Leftrightarrow \hat{f} = \left[\sum_{i=1}^{n} D_i^T V_i^{-1}(Y_i - \mu_i) - \lambda G\right]^{-1}$$

Kemudian menggunakan estimasi parameter ke t untuk memperbaharui $\hat{f}$ dalam persamaan sebagai berikut :

$$\hat{f}^{t+1} = \hat{f}^{t} + \left[\sum_{i=1}^{n} D_i^T V_i^{-1}(Y_i - \mu_i) - \lambda G\right]^{-1}\left[\sum_{i=1}^{n} D_i^T V_i^{-1}(Y_i - \mu_i) - \lambda G \hat{f}^{t}\right]$$

B. Deskripsi Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil di Rumah Sakit Unhas Makassar. Populasi dari penelitian ini adalah pasien DBD yang pernah menjalani rawat inap di Rumah Sakit Unhas Makassar. Sampel dari penelitian ini berasal dari data rekam medis pasien DBD periode bulan Januari sampai bulan Maret 2018 sebanyak 58 sampel. Peubah respon (Y) yaitu lama kesembuhan pasien DBD (hari) dan peubah bebas (Xi) yaitu umur (tahun), suhu tubuh (℃), trombosit ($\times 10^3 \mu L$), dan hemoglobin (g/dL). Statistika deskriptif dari variabel-variabel penelitian diolah dengan bantuan program Minitab 17.

Tabel 4.1 Statistika deskriptif data pasien Demam Berdarah Dengue (DBD)

Variabel	**Ringkasan Statistik**			
	Min	Max	Mean	StDev
Lama Kesembuhan (Y)	2,0	9,0	4,776	1,511
Suhu (S)	36,240	38,733	37,050	0,566
Umur (U)	1,0	66,0	22,22	15,60
Trombosit (PLT)	32,3	421,8	147,5	88,2
Hemoglobin (HB)	10,633	15,967	13,636	1,463

C. Penentuan Komponen Parametrik dan Komponen Nonparametrik

Tahap awal sebelum melakukan pemodelan regresi dalam hal ini memodelkan kasus Demam Berdarah Dengue di Makassar adalah menentukan variabel parametrik dan nonparametrik dengan melakukan pengujian. Untuk mengetahui data tersebut dalam kelompok parametrik atau nonparametrik, terlebih dahulu akan dilakukan uji normalitas terhadap datadengan cara melihat plot dari data tersebut dengan menggunakan hipotesis sebagai berikut :

H_0: $\mu = 0$ (normal)

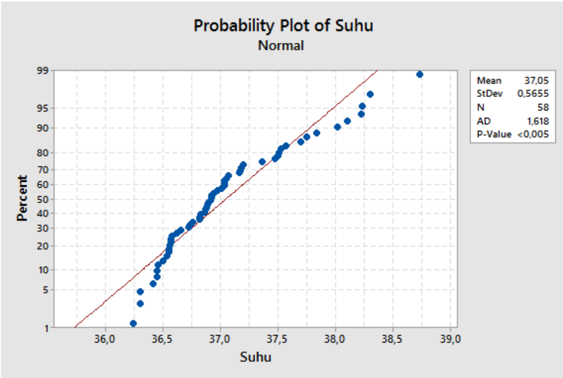
H_1: $\mu \neq 0$ (tidak normal)

Dalam hal pengujian hipotesis ini, kriteria untuk menolak atau menerima H_0 berdasarkan *p-value* atau nilai signifikansi uji yang dinyatakan sebagai berikut :

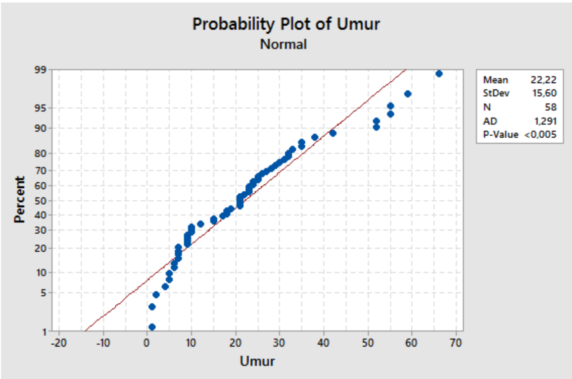
Jika *p-value*$< \alpha = 0.05$, maka H_0 ditolak

Jika *p-value*$> \alpha = 0.05$, maka H_0 diterima

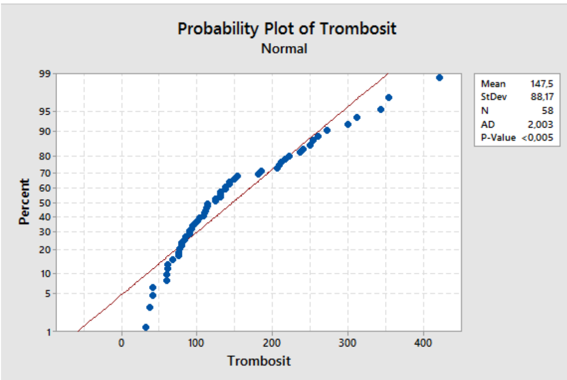
Hasil uji normalitas untuk masing-masing variabel prediktor adalah sebagai berikut :



Gambar 4.1 Plot Uji Normalitas pada Suhu Tubuh

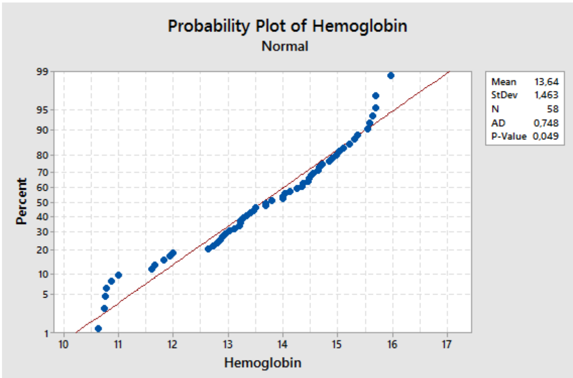


Gambar 4.2 Plot Uji Normalitas pada Umur



Gambar 4.3 Plot Uji Normalitas pada Trombosit

Dari gambar 4.1, 4.2 dan 4.3 diatas menunjukkan bahwa *p-value* <0.05 maka tolak H_0 dan terima H_1 yang berarti bahwa data tidak menyebar normal.



Gambar 4.4 *Scatter Plot* antara Lama Kesembuhan (Y) dengan Hemoglobin

Ganbar 4.1 menunjukkan bahwa *p-value* > 0.05 maka tolak H_0 dan H_1 yang berarti bahwa data menyebar normal.

Dengan demikian data pasien Demam Berdarah Dengue di Kota Makassar 2018 dapat didekati dengan regresi semiparametrik spline, dimana terdapat 1 variabel prediktor yang merupakan komponen parametrik yaitu Hemoglobin dan tiga variabel prediktor yang merupakan komponen nonparametrik yaitu Suhu, Umur dan Trombosit.

Tabel 4.2 Komponen parametrik dan komponen nonparametrik regresi semiparametrik spline

Variabel	Komponen
Usia ($\boldsymbol{Z_1}$)	Nonparametrik
Suhu ($\boldsymbol{Z_2}$)	Nonparametrik
Trombosit ($\boldsymbol{Z_3}$)	Nonparametrik
Hemoglobin ($\boldsymbol{X}$)	Parametrik

D. Model Regresi Semiparametrik Spline

Penentuan model regresi semiparametrik *spline* dipengaruhi oleh pemilihan titik knot optimal. Titik knot merupakan titik perpaduan bersama yang memperlihatkan terjadinya perubahan perilaku dari fungsi *spline* pada interval-interval yang berbeda sehingga kurva yang terbentuk tersegmen pada titik tersebut. Pemilihan titik knot optimal dapat dilakukan dengan melihat nilai *Generalized Cross Validation* (GCV) yang minimum. Berdasarkan penentuan komponen parametrik dan komponen nonparametrik, model regresi semiparametrik *spline* dapat dituliskan dalam persamaan berikut :

$$Y_i = \beta_0 + \beta_1 X_{1n} + \beta_2 Z_{1n} + \sum_{j=1}^{m} \beta_j \left(Z_{1n} - k_j\right) + \beta_3 Z_{2n} + \sum_{j=1}^{m} \beta_j \left(Z_{1n} - k_j\right) + \beta_4 Z_{3n} + \sum_{j=1}^{K} \beta_j \left(Z_{1n} - k_j\right) + \varepsilon_i \qquad (4.1)$$

Dengan

$$X = \begin{bmatrix} 1 & x_{11} & z_{11} & z_{21} & z_{31} \\ 1 & x_{12} & z_{12} & z_{22} & z_{32} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1(58)} & z_{1(58)} & z_{2(58)} & z_{3(58)} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$$

$$Z = \begin{bmatrix} (z_{11} - k_1) & (z_{21} - k_2) & (z_{31} - k_3) \\ (z_{12} - k_1) & (z_{22} - k_2) & (z_{32} - k_3) \\ \vdots & \vdots & \vdots \\ (z_{1(58)} - k_1) & (z_{2(58)} - k_2) & (z_{3(58)} - k_3) \end{bmatrix},$$

Dimana X merupakan komponen parametrik, Z adalah komponen nonparametrik, k adalah titik knot.

E. Pemilihan Titik Knot Optimal Regresi Semiparametrik *Spline*

Titik knot merupakan titik perubahan perilaku data pada sub-sub interval tertentu. Titik knot yang dicobakan pada penelitian ini sampai tiga titik knot ($k = 1,2,3$), bertujuan agar memudahkan peneliti dalam melakukan interpretasi. Untuk mendapatkan titik knot yang optimal, digunakan metode *Generalize Cross Validation* (GCV). Nilai GCV yang paling minimum diantara ketiga titik knot merupakan titik knot yang optimal.

1. Pemilihan Titik Knot dengan Satu Titik Knot

Estimasi model regresi semiparametrik spline dengan satu titik knot pada data pasien demam berdarah dengue di Rumah Sakit UNHAS Makassar adalah sebagai berikut :

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_{1n} + \hat{\beta}_2 z_{1n} + \hat{\beta}_3 (z_{1n} - k_1) + \hat{\beta}_4 (z_{2n} - k_2) + \hat{\beta}_5 (z_{3n} - k_3)$$

Tabel 4.3 menunjukkan sepuluh nilai GCV yang berada disekitar nilai GCV paling minimum untuk model regresi semiparametrik spline satu knot.

Tabel 4.3 Nilai GCV Satu Titik Knot

Z_1	Z_2	Z_3	GCV
38.22448980	342.26020408	14.87823129	268.3952118
38.27537415	350.20918367	14.98707483	267.6062819
38.32625850	358.15816327	15.09591837	267.2516789
38.37714286	366.10714286	15.20476190	267.2301903
38.42802721	374.05612245	15.31360544	266.9158093
38.47891156	382.00510204	15.42244898	266.947996
38.52979592	389.95408163	15.53129252	267.1483543
38.58068027	397.90306122	15.64013605	261.2325882
38.63156463	405.85204082	15.74897959	260.4409289
38.68244898	**413.80102041**	**15.85782313**	**244.8255613**

Berdasarkan Tabel 4.3 diketahui bahwa nilai GCV minimum untuk model regresi semiparametrik spline dengan satu titik knot adalah sebesar 244.8255613. Nilai tersebut diperoleh dari satu titik knot optimal pada setiap variabel prediktor. Titik knot optimal untuk variabel umur (Z_1) berada pada titik knot 38.68244898, variabel suhu (Z_2) berada pada titik knot 413.80102041, dan variabel trombosit (Z_3) berada pada titik knot 15.85782313.

2. Pemilihan Titik Knot dengan Dua Titik Knot

Setelah dilakukan pemilihan titik knot dengan satu titik knot, selanjutnya dilakukan pemilihan titik knot optimal menggunakan dua titik knot pada setiap variable nonparametrik. Berikut merupakan model regresi semiparametrik spline dari lama kesembuhan pasien demam berdarah dengue di Rumah Sakit UNHAS Makassar dengan dua titik knot.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_{1n} + \hat{\beta}_2 z_{1n} + \hat{\beta}_3 (z_{1n} - k_1) + \hat{\beta}_4 (z_{1n} - k_1) + \hat{\beta}_5 z_{2n} +$$
$$\hat{\beta}_6 (z_{2n} - k_2) + \hat{\beta}_7 (z_{2n} - k_2) + \hat{\beta}_8 z_{3n} + \hat{\beta}_9 (z_{3n} - k_3) + \hat{\beta}_{12} (z_{3n} - k_3)$$

Tabel 4.4 menunjukkan sepuluh nilai GCV yang berada disekitar nilai GCV paling minimum untuk model regresi semiparametrik spline dua knot.

Tabel 4.4 Nilai GCV Dua Titik Knot

$\boldsymbol{Z_1}$	$\boldsymbol{Z_2}$	$\boldsymbol{Z_3}$	GCV
38.52979592	389.95408163	15.53129252	266.86969483
38.58068027	397.90306122	15.64013605	
38.52979592	389.95408163	15.53129252	266.76715638
38.63156463	405.85204082	15.74897959	
38.52979592	389.95408163	15.53129252	267.03225635
38.68244898	413.80102041	15.85782313	
38.52979592	389.95408163	15.53129252	267.14835426
38.73333333	421.75	15.96666667	
38.58068027	397.90306122	15.64013605	268.11085749
38.63156463	405.85204082	15.74897959	
38.58068027	397.90306122	15.64013605	268.96409223
38.68244898	413.80102041	15.85782313	
38.58068027	397.90306122	15.64013605	261.23258819
38.73333333	421.75	15.96666667	
38.63156463	405.85204082	15.74897959	260.44984795
38.68244898	413.80102041	15.85782313	
38.63156463	405.85204082	15.74897959	260.44092891
38.73333333	421.75	15.96666667	
38.68244898	**413.80102041**	**15.85782313**	**244.82556133**
38.73333333	**421.75**	**15.96666667**	

Berdasarkan Tabel 4.4 diketahui bahwa nilai GCV minimum untuk model regresi semiparametrik spline dengan dua titik knot adalah sebesar 244.82556133. Nilai tersebut diperoleh dari dua titik knot optimal pada setiap variabel prediktor. Titik knot optimal untuk variabel umur (Z_1) berada pada titik knot 38.68244898 dan 38.73333333, variabel suhu (Z_2) berada pada titik knot 413.80102041 dan 421.75 dan variabel trombosit (Z_3) berada pada titik knot 15.85782313 dan 15.96666667.

3. Pemiliihan Titik Knot dengan Tiga Titik Knot

Estimasi model regresi semiparametrik spline dengan tiga titik knot pada data pasien demam berdarah dengue di Rumah Sakit UNHAS Makassar adalah sebagai berikut :

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_{1n} + \hat{\beta}_2 z_{1n} + \hat{\beta}_3 (z_{1n} - k_1) + \hat{\beta}_4 (z_{1n} - k_1) + \hat{\beta}_5 (z_{1n} - k_1) + \hat{\beta}_6 z_{2n} + \hat{\beta}_7 (z_{2n} - k_2) + \hat{\beta}_8 (z_{2n} - k_2) + \hat{\beta}_9 (z_{2n} - k_2) + \hat{\beta}_{10} z_{3n} + \hat{\beta}_{11} (z_{3n} - k_3) + \hat{\beta}_{12} (z_{3n} - k_3) + \hat{\beta}_{13} (z_{3n} - k_3)$$

Tabel 4.5 Nilai GCV Tiga Titik Knot

$\boldsymbol{Z_1}$	$\boldsymbol{Z_2}$	$\boldsymbol{Z_3}$	GCV
38.07183673	318.41326531	14.55170068	
38.22448980	342.26020408	14.87823129	271.44658424
38.52979592	389.95408163	15.53129252	
38.07183673	318.41326531	14.55170068	
38.22448980	342.26020408	14.87823129	272.03787270
38.58068027	397.90306122	15.64013605	
38.07183673	318.41326531	14.55170068	
38.22448980	342.26020408	14.87823129	266.79574691
38.63156463	405.85204082	15.74897959	
38.07183673	318.41326531	14.55170068	
38.22448980	342.26020408	14.87823129	256.21704862
38.68244898	413.80102041	15.85782313	
38.07183673	**318.41326531**	**14.55170068**	
38.27537415	**350.20918367**	**14.98707483**	**221.67745153**
38.32625850	**358.15816327**	**15.09591837**	
38.07183673	318.41326531	14.55170068	
38.27537415	350.20918367	14.98707483	221.70268020
38.37714286	366.10714286	15.20476190	
38.07183673	318.41326531	14.55170068	230.51608266
38.27537415	350.20918367	14.98707483	

38.42802721	374.05612245	15.31360544	
38.07183673	318.41326531	14.55170068	
38.27537415	350.20918367	14.98707483	229.66947089
38.47891156	382.00510204	15.42244898	
38.07183673	318.41326531	14.55170068	
38.27537415	350.20918367	14.98707483	273.55980153
38.52979592	389.95408163	15.53129252	
38.07183673	318.41326531	14.55170068	
38.27537415	350.20918367	14.98707483	274.83043678
38.58068027	397.90306122	15.64013605	

Berdasarkan Tabel 4.5 diketahui sepuluh nilai GCV yang berada disekitar nilai GCV paling minimum untuk model regresi semiparametrik spline tiga titik knot. Pada Tabel 4.5 diketahui bahwa nilai GCV minimum untuk model regresi semiparametrik spline dengan tiga titik knot adalah sebesar 221.67745153. Nilai tersebut diperoleh dari tiga titik knot optimal pada setiap variabel prediktor. Titik knot optimal untuk variabel umur (Z_1) berada pada titik knot 38.07183673, 38.27537415 dan 38.32625850, variabel suhu (Z_2) berada pada titik knot 318.41326531, 350.20918367 dan 358.15816327 dan variabel trombosit (Z_3) berada pada titik knot 14.55170068, 14.98707483 dan 15.09591837.

4. Pemilihan Titik Knot Terbaik

Titik knot terbaik merupakan titik knot yang mempunyai nilai GCV dan MSE minimum. Berikut merupakan perbandingan nilai GCV dan MSE minimum yang diperoleh pada satu titik knot, dua titik knot, dan tiga titik knot yang ditunjukkan pada Tabel 4.6.

Tabel 4.6 Perbandingan Nilai GCV dan MSE

Model	GCV	MSE
1 Titik Knot	244.82556133	204.4337
2 Titik Knot	244.82556133	204.4337
3 Titik Knot	**221.67745153**	**199.1032**

Berdasarkan kriteria pemilihan model terbaik diketahui bahwa nilai GCV dan MSE paling minimum dihasilkan oleh model regresi nonparametrik spline dengan tiga titik knot.

F. Pengujian Parameter Model

Setelah didapatkan model regresi semiparametrik spline terbaik, kemudian dilakukan pengujian signifikansi parameter model regresi semiparametrik spline.

1. Uji Serentak

Dilakukan uji signifikasi parameter regresi semiparametrik spline secara serentak dengan uji F dengan hipotesis sebagai berikut :

$$H_0: \beta_1 = \cdots = \beta_{13} = 0$$

H_1: tidak semua koefisien regresi bernilai nol

Berikut merupakan analisis ragam dari model regresi semiparametrik yang disajikan pada Tabel 4.7.

Tabel 4.7 Uji serentak estimasi model regresi semiparametrik spline

Sumber Keragaman	**Derajat Bebas**	**Jumlah Kuadrat (JK)**	**Rataan Jumlah Kuadrat (RJK)**	F_{hitung}	F_{tabel} $\alpha = 5\%$
Regresi	12	11885,33	990,4442	22,4788	1,97
Residual	45	1982,754	44,0612		
Total	57	13868,09			

Dengan taraf nyata $\alpha = 5\%$ diperoleh kesimpulan bahwa $F_{hitung} \geq F_{tabel}$ yaitu 22,4788$\geq$ 1,97 maka tolak H_0 yang mengindikasikan bahwa tidak semua koefisien regresi bernilai nol atau dengan kata lain terdapat pengaruh yang signifikan secara bersama-sama antara variabel bebas terhadap variabel terikat, sehingga model signifikan.

2. Uji Individu

Uji parameter secara parsial (secara individu) menggunakan pendekatan GEE dapat dilihat berdasarkan nilai signifikansi atau *p-value* dari uji Wald.

Hasil pengujian tersebut disajikan dalam bentuk Tabel 4.8

Tabel 4.8 Uji individu estimasi model regresi semiparametrik spline

Variabel	Parameter	Estimasi	$p - value$	Keterangan
	β_0	332,588	0,003	Signifikan
$\boldsymbol{X}$	β_1	0,227	0,000	Signifikan
$\boldsymbol{Z_1}$	β_2	0,012	0,058	Tidak signifikan
	β_3	-7,827	0,617	Tidak signifikan
	β_4	0,379	0,840	Tidak signifikan
	β_5	0,008	0,993	Tidak signifikan
$\boldsymbol{Z_2}$	β_6	-0,056	0,503	Tidak signifikan
	β_7	0,004	0,438	Tidak signifikan
	β_8	0,119	0,010	Signifikan
	β_9	0,002	0,819	Tidak signifikan
$\boldsymbol{Z_3}$	β_{10}	-0,004	0,000	Signifikan
	β_{11}	2,974	0,682	Tidak signifikan
	β_{12}	-8,745	0,012	Signifikan
	β_{13}	-0,41	0,757	Tidak signifikan

Tabel 4.8 menjelaskan bahwa dari keempat variabel prediktor, dua variabel prediktor mempunyai parameter yang signifikan terhadap model karena memiliki *p-value* kurang dari 5%, sehingga variabel hemoglobin (X) dan trombosit (Z_3)

berpengaruh secara signifikan terhadap lama kesembuhan pasien di Rumah Sakit Unhas Makassar.

G. Pengujian Residual Model

Residual (goodness of fit) dari suatu model regresi harus memenuhi asumsi $IIDN(0, \sigma^2)$ yaitu identik, independen, dan berdistribusi normal dengan mean nol dan variansi σ^2.

1. Uji Asumsi Homogenitas

Uji asumsi homogenitas bertujuan untuk melihat apakah kelompok data yang digunakan memiliki varians yang relatif sama (homogen). Uji asumsi homogenitas dapat dilakukan dengan menggunakan uji Glejser. Dengan bantuan program IBM SPSS Statistics 22 diperoleh nilai F_{hit} sebesar 0,000 pada taraf $\alpha = 5\%$, F_{tabel} sebesar 2,55; diperoleh bahwa nilai F_{hit} lebih kecil dari F_{tabel} yang mengindisikan terima H_0. Maka dapat disimpulkan bahwa semua variabel tidak berpengaruh signifikan terhadap nilai mutlak residual. Hal tersebut membuktikan bahwa varians residual memenuhi asumsi homokedastisitas atau dengan kata lain tidak terjadi heterokedastisitas.

2. Uji Asumsi Independen

Uji asumsi independen dilakukan untuk mengetahui apakah terdapat korelasi pada residual. Asumsi residual independen dapat dilakukan dengan menggunakan uji d Durbin-Watson, dengan bantuan IBM SPSS Statistics 22 diperoleh nilai d Durbin-Watson sebesar 2,027. Selanjutnya nilai d Durbin-Watson tersebut akan dibandingkan dengan nilai tabel signifikasi $\alpha = 5\%$, dengan jumlah sampel sebanyak 58 (T=58), satu variabel dependen dan empat variabel independen (k=5). Dari tabel d Durbin-Watson dengan $\alpha = 5\%$. T=58 dan k=5 diperoleh nilai d_L dan d_U secara berturut-turut yaitu 1,3953 dan 1,7673

$$4 - d_L = 4 - 1{,}3953 = 2{,}6047$$

$$4 - d_U = 4 - 1{,}7673 = 2{,}2327$$

Karena nilai d Durbin-Watson terletak diantara nilai d_U dan $4 - d_U$ maka terima H_0 yang mengindikasikan bahwa tidak terdapat autokorelasi positif ataupun negatif pada residual.

3. Uji Asumsi Normal

Uji normalitas dilakukan untuk melihat apakah residual mengikuti distribusi normal atau tidak. Pengujian asumsi normalitas dapat dilakukan dengan melakukan uji Anderson-Darling, dengan bantuan Minitab 17 diperoleh nilai Anderson-Darling sebesar 0,384 dan $p - value$ sebesar 0,384 pada taraf nyata 5%. Karena $p - value$ lebih besar dari 0,05, maka terima H_0 yang mengindikasikan bahwa residual model memenuhi asumsi distribusi normal.

Dengan demikian dapat disimpulkan bahwa residual dari model regresi semiparametrik spline linear dengan tiga titik knot memenuhi asumsi $IIDN(0, \sigma^2)$ yaitu identik, independen, dan berdistribusi normal.

H. Koefisien Determinasi

Nilai koefisien determinasi (R^2) menunjukkan seberapa besar kebaikan model regresi dalam menjelaskan keragaman lama kesembuhan pasien demam berdarah dengue di Rumah Sakit Unhas Makassar

$$R^2 = \frac{SS_{regresi}}{SS_{total}} \times 100\%$$

$$= \frac{11885,33}{13868,09} \times 100\%$$

$$= 85,7\%$$

Berdasarkan perhitungan didapatkan nilai R^2 sebesar 85,7%. Hal ini berarti model regresi semiparametrik spline yang didapatkan mampu menjelaskan keragaman lama kesembuhan pasien demam berdarah dengue di Rumah Sakit Unhas Makassar. Nilai tersebut mendekati 100%, sehingga model sudah cukup baik.

I. Interpretasi Model Regresi Semiparametrik Spline dengan Tiga Titik Knot

Model terbaik lama kesembuhan pasien Demam Berdarah Dengue (DBD) di Rumah Sakit UNHAS Makassar menggunakan regresi semiparametrik spline adalah model semiparametrik spline dengan tiga titik knot, adapun estimasi model yang diperoleh adalah sebagai berikut :

$$\hat{y} = 332,588 + 0,227x - 0,004z_3 + 2,974(z_3 - 14,552) -$$

$$8,745(z_3 - 14,987) - 0,41(z_3 - 15,096)$$

Berdasarkan model tersebut, maka dapat diinterpretasikan sebagai berikut.

1. Apabila variabel hemoglobin (X) dinggap konstan, maka pengaruh trombosit (Z_3) terhadap lama kesembuhan pasien adalah

$$\hat{y} = 332{,}588 - 0{,}004z_3 + 2{,}974(z_3 - 14{,}552) -$$

$$8{,}745(z_3 - 14{,}987) - 0{,}41(z_3 - 15{,}096)$$

$$= \begin{cases} 332{,}588 - 0{,}004z_3 & ; z_3 < 14{,}552 \\ 289{,}310 + 2{,}97z_3 & ; 14{,}552 \leq z_3 < 14{,}987 \\ 463{,}649 - 8{,}749z_3 & ; 14{,}987 \leq z_3 < 15{,}096 \\ 338{,}777 - 0{,}414z_3 & ;\ z_3 \geq 15{,}096 \end{cases}$$

Ketika jumlah trombosit naik sebesar 1$(\times\ 10^3\mu L)$ pada saat jumlah trombosit kurang dari 14,552 maka y akan mengalami penurunan sebesar 0,004. Apabila jumlah trombosit berada diantara 14,552 dan 14,987 maka y akan mengalami kenaikan 2,97. Apabila jumlah trombosit berada diantara 14,987 dan 15,096, maka y akan mengalami penurunan sebesar 8,749. Dan apabila kadar trombosit lebih dari 15,096, maka y akan mengalami penurunan sebesar 0,003. Koefisien bernilai negatif artinya terjadi hubungan negatif antara trombosit dengan y yang mengindikasikan bahwa apabila kadar trombosit meningkat menyebabkan kesembuhan pasien cenderung semakin cepat. Begitupula sebaliknya, koefisien bernilai positif artinya terjadi hubungan positif antara trombosit dengan y yang mengindikasikan bahwa apabila kadar trombosit meningkat menyebabkan kesembuhan pasien cenderung semakin lambat.

2. Apabila trombosit (Z_3) dianggap konstan, maka interpretasi terhadap variabel hemoglobin adalah apabila hemoglobin mengalami kenaikan 1(g/dL) maka y akan mengalami kenaikan sebesar 0,227. Koefisien bernilai positif artinya terjadi hubungan positif antara hemoglobin dengan y. Hal ini mengindikasikan bahwa apabila jumlah hemoglobin meningkat maka berakibat pada jenjang waktu yang lebih lama pada kesembuhan pasien.

J. Pembahasan

Berdasarkan beberapa penelitian sebelumnya, yang dilakukan oleh Utami (2014) yaitu "Pemodelan Regresi Nonparametrik Pada Data Longitudinal Berdasarkan Estimator Polinomial Lokal Kernel" studi kasus Demam Berdarah Dengue (DBD) yang menghasilkan bentuk estimasi model regresi nonparametrik pada data longitudinal berdasarkan estimator polinomial lokal kernel adalah $\hat{\eta}(t_{ij}) = x_{ij}^T(X^TK_hX)^{-1}X^TK_hy$. Sedangkan pada penelitian ini, "Model Regresi Semiparametrik *Spline* Untuk Data Longitudinal Pada Kasus Penderita DBD" namun menggunakan pendekatan GEE yang menghasilkan bentuk estimator yaitu $\hat{\beta} = [\sum_{i=1}^{n} X_i^T V_i^{-1} X_i]^{-1}[\sum_{i=1}^{n} X_i^T V_i^{-1} Y_i]$.

Penelitian yang dilakukan oleh Purhadi (2012) yaitu "Analisis *Survival* Faktor-faktor yang Mempengaruhi Laju Kesembuhan Pasien Penderita Demam Berdarah Dengue (DBD) di RSU Haji Surabaya dengan Regresi *Cox*" menggunakan variabel prediktor usia, jenis kelamin, hemoglobin, leukosit, hematokrit, dan trombosit yang menghasilkan bahwa faktor-faktor yang mempengaruhi laju kesembuhan pasien adalah variabel usia dan trombosit.

Sedangkan hasil penelitian Model Regresi Semiparametrik *Spline* Untuk Data Longitudinal Pada Kasus Penderita DBD menghasilkan variabel prediktor yang paling signifikan terhadap laju kesemubuhan pasien adalah hemoglobin dan trombosit.

Penelitian yang dilakukan oleh Yani (2017) yaitu "Aplikasi Model Regresi Semiparametrik *Spline Truncated*" studi kasus penderita Demam Berdarah Dengue (DBD) diperoleh :

$$\hat{Y}_i = -0{,}55021 - 0{,}01136HCT_i + 0{,}01648(HCT_i - 39{,}6) + 0{,}28815S_i - 0{,}0475U_i - 0{,}01104PLT_i$$

dengan nilai GCV minimum sevesar 0,03553, nilai MSE sebesar 0,02969, nilai koefisien determinasi sebesar 98,91% dengan enam parameter. Sedangkan pada penelitian ini, "Model Regresi Semiparametrik *Spline* Untuk Data Longitudinal Pada Kasus Penderita DBD" diperoleh

$$\hat{y} = 332{,}588 + 0{,}227x - 0{,}004z_3 + 2{,}974(z_3 - 14{,}552) - 8{,}745(z_3 - 14{,}987) - 0{,}41(z_3 - 15{,}096)$$

dengan nilai GCV minimum 221.67745153**,** nilai MSE sebesar 199.1032**,** nilai koefisien determinasi sebesar 85,7%.

BAB V

SIMPULAN DAN SARAN

A. Simpulan

Estimasi model regresi semiparametrik dengan kriteria nilai GCV minimum pada model regresi semiparametrik *spline* dengan tiga titik knot, diperoleh model sebagai berikut:

$$\hat{y} = 332,588 + 0,227x - 0,004z_3 + 2,974(z_3 - 14,552) - 8,745(z_3 - 14,987) - 0,41(z_3 - 15,096)$$

Dengan nilai GCV minimum 221.67745153**,** nilai MSE sebesar 199.1032 yang dicapai pada titik knot 14,552; 14,987; 15,096, memiliki koefisien determinasi sebesar 85,7% keragaman lama kesembuhan pasien Demam Berdarah Dengue (DBD) yang menjalani rawat inap di Rumah Sakit Unhas Makassar.

B. Saran

Dalam penelitian ini dibahas model regresi semiparametrik untuk satu variabel pada komponen parametriknya sehingga penelitian lebih lanjut dapat dilakukan untuk lebih dari satu varibel komponen parametrik.

DAFTAR PUSTAKA

Abdy, M. 2009. Regresi Semiparametrik dengan Pendekatan *Generalized Estimating Equation (GEE). Jurnal Matematika, Statistika dan Komputasi*, Vol. 5, No.2, 66-75.

Adawiyah, R. 2018. Model Regresi Nonparametrik dengan Pendekatan Spline (Studi Kasus : Berat Badan Lahir Rendah di Rumah Sakit Ibu dan Anak Siti Fatimah Makassar). *Skripsi.* Makassar : Program Studi Matematika Universitas Negeri Makassar.

Danardono. 2018. *Analisis Data Longitudinal*. Yogyakarta : UGM Press.

Gusti, O.W. 2011. Regresi Semiparametrik Spline Dalam Memodelkan Hasil UNAS SMAN 1 Sekaran Lamongan. *Skripsi*. Malang : Program Studi Matematika Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Handayani, L dan Putera, F.H.A. 2016. Model Data Longitudinal dengan Pendekatan *Generalized Estimating Equation* pada Struktur Korelasi *Exchangeable, Auto-Regressive,* dan *Unstructured. Jurnal Statistika Universitas Tadulako.*

Handayanti, K. 2015. Kajian Metode Generalized Estimating Equation (GEE) Dalam Pendugaan Parameter Model Regresi Multilevel. *Jurnal Matematika Murni dan Terapan*, Vol.12, No.2, 500-510.

Ibrahim, N.A. dan Suliadi. 2009. Nonparametric Regression for Correlated Data. *Article in WSEAS Transaction on Mathemathics.* Issue 7, Vol.8, ISSN:1109-2769.

Laome, L. 2009. Model Regresi Semiparametrik Spline Untuk Data Longitudinal pada Kasus Kadar CD4 Penderita HIV. *Jurnal Matematika Murni dan Terapan,* Vol.13 No.2,189-194.

Latra, I. N. 2013. Analisis *Survival* dengan Model Regresi *Cox Weibull* pada Penderita Demam Berdarah Dengue (DBD) di Rumah Sakit Haji Sukolilo Surabaya. *Jurnal Sains dan Seni POMITS,* Vol.2, No.2, 2337-3520.

Nirmala, F. 2013. Aplikasi GLMM pada Data Longitudinal Kadar Trombosit Demam Berdarah Dengue. *Jurnal Biometrika dan Kependudukan,* Vol. 2, No. 2 Desember 2013: 131–139.

Poerwanto, B dan Budiantara, I. N. 2014. Estimasi Kurva Regresi Semiparametrik Spline Untuk Data Longitudinal. *Prosiding Seminar Nasional Matematika*, Universitas Udayana, 6 November 2014

Purhadi. 2012. Analisis *Survival* Faktor-faktor yang Mempengaruhi Laju Kesembuhan Pasien Penderita Demam Berdarah Dengue (DBD) di RSU Haji Surabaya dengan Regresi *Cox. Jurnal Sains dan Seni ITS* , Vol.1, No.1, 2301-928X.

Utami, T. W. 2014. Pemodelan Regresi Nonparametrik pada Data Longitudinal Berdasarkan Estimasi Polonomial Lokal Kernel. *Jurnal Statistika* Vol.5 No.2, 602-610

Yani, N. W. M. N. 2017. Aplikasi Model Regresi Semiparametrik Spline Truncated *Jurnal Matematika* Vol.6 No.1, 65-73.

RIWAYAT HIDUP

Mustati'atul Waidah Maksum, lahir di Sinjai, Kecamatan Sinjai Utara, Kabupaten Sinjai pada tanggal 17 November 1996 sebagai anak kelima dari pasangan Maksum dan Cahaya Djunaid . Penulis memulai jenjang pendidikan sekolah dasar di SDN 4 Balangnipa pada tahun 2002 dan tamat tahun 2008. Pada tahun yang sama penulis melanjutkan pendidikan menengah pertama di SMPN 1 Sinjai dan tamat tahun 2011. Kemudian melanjutkan studi di SMA Negeri 1 Sinjai pada tahun 2011 dan tamat tahun 2014. Penulis melanjutkan studi ke jenjang perguruan tinggi pada tahun 2014 Program studi Teknik di salah satu Universitas Swasta dan kemudian melanjutkan ke perguruan tinggi negeri pada tahun 2015 di Program Studi Matematika Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Makassar melalui jalur SBMPTN dan menyelesaikan studi S1 pada tahun 2019.